Computational Economics
https://doi.org/10.1007/s10614-026-11407-7

Interpretable LIM-GRLST Hybrid Deep Learning Approach for Financial Cryptocurrency Trends Prediction

Wajeeha Badar[1] · Shabana Ramzan[1] · Ali Raza[2] · David Ramirez Alcocer[3,4,5,6] · Gerardo Mendez Mezquita[3,6,7,8] · Imran Ashraf[9]

Received: 17 December 2025 / Accepted: 23 July 2026
© The Author(s) 2026

Abstract
Bitcoin and other cryptocurrency markets are volatile by nature, which provides an opportunity as well as a threat to an investor and researcher. The correct forecasting of prices is an important factor in making informed decisions when in such dynamic conditions. The paper presents a explainable LIM-GRLST framework of Bitcoin price forecasting on a one-step-ahead basis with deep learning. The baseline LSTM and GRU models were trained initially and then hybrid models that combine the complementary advantages of the two models were trained. A weighted hybrid also exhibited relatively small improvements, whereas the neural hybrid produced the highest results (MAE = 0.0209, MSE = 0.0008, RMSE = 0.0279, R 2 = 0.9897), which were tested on an unknown test set. Standard post-hoc methods, permutation importance and LIME, were used to provide the feature-level explainability, indicating closing, high, and low prices as the most important predictors. The article focuses on enhancing engineering within frameworks of hybrid and hyperparameter optimization over a novel methodology, and is shown to be applicable in practice to cryptocurrency forecasting.

Keywords Cryptocurrency · Time series forecasting · Deep learning · Neural hybridization · Permutation · Explainable artificial intelligence

1 Introduction

Cryptocurrency market and Bitcoin, Litecoin, and Ethereum market in particular are extremely volatile, volatile and are prone to numerous world economic and regulatory factors. It is under this uncertainty that there is need to have accurate price prediction in order to help in decision-making in trading and managing portfolios and risk management. Deterministic or probabilistic predictive models that are able to

Extended author information available on the last page of the article

Published online: 06 August 2026

capture the complex time-dependent and nonlinear behavior of price data are needed by the investors and researchers to reduce risk and maximize returns.

In the recent research, the increasing number of studies has applied more deep learning (DL) structures to improve the prediction of the prices of cryptocurrencies. Indicatively, developed predictive models using long short term memory (LSTM) and Gated Recurrent Unit (GRU) networks to forecast Bitcoin, Ethereum and Litecoin, and demonstrate superior forecasting models compared to traditional models (Kaur et al., 2025b). Similarly, bi-directional LSTM (Bi-LSTM) is more useful in capturing bi-directional time dependencies compared to unidirectional LSTM when predicting cryptocurrencies (Seabe et al., 2023). Additional encode-rdecoder LSTM and GRU models have been explored and are also competitive in stock and cryptocurrency prediction (Das et al., 2024).

However, the majority of the research is more concerned with predictive performance and is not systematic in the feature-level interpretability. One such case is that a deep learning model was built predicting multi-currency, yet lacked any explainability mechanisms to predict it using Bi-LSTM and GRU with zero explanations behind it (Aljadani, 2022). A modification of this work was conducted by Birim et al. who used Bi-LSTM and Bi-GRU with the emphasis on the accuracy with minimum post-hoc interpretations again (Birim et al., 2022). The comparative studies of Lian demonstrated that hybrid models founded on LSTM and GRU are more effective than single ones, but underutilize the tools of interpretability (Lian, 2024) and Murray et al. have proven the effectiveness of hybrid models without implementing the tools of explainability in their entirety. The fact that more and more people use DL and hybrid models to predict the price of cryptocurrencies is also supported by a number of survey and practicum studies (Murray et al., 2023). A systematic review of deep learning applications in cryptocurrency trading that needs accuracy and interpretability (Ataei et al., 2025). Similarly, the development of hybrid DL models but further reported that plain prediction could not be applied in real-life decision support without an explanation ability (Amberkhani et al., 2025).

Other recent studies confirm these results. Having compared CNN, GRU, and LSTM models, it was found that GRU was quite helpful to predict Bitcoin on a daily basis (Alvi et al., 2024). LSTM is routinely more accurate with prediction than RNN, GRU, and Bi-LSTM, as well as, in accuracy as compared to Bi-LSTM alone (Nair et al., 2023). Hybrids of CNN-GRU were also explored which have higher combined accuracy-complexity in predicting BTC and ETH simultaneously (Mazinani et al., 2025). Analyzed hybrid models of cryptocurrency and commodity price prediction and the benefits of including CNN, LSTM, and GRU layers (Gur, 2024). Based on historical financial data made it clear that hybrid LSTM-GRU models are superior to single architectures in predicting Bitcoin prices (Hussein & Abdulazeez, 2024).

Although there has been a recent progress in cryptocurrency forecasting, most of the available literature mainly concentrates on predictive performance, and typically adopts conventional deep learning architectures without strategically adding explainability. Moreover, little research has been done on hybrid architectures along with thorough hyperparameter optimization and statistically justified evaluation (e.g., DieboldMariano test) to guarantee coherent and significant performance gains.

To fill these gaps, this paper suggests a hybrid LSTMGRU model to predict one-day-ahead Bitcoin closing price. The model uses past 30 daily observations sequences to predict the closing price of the next day. To maximize the predictive accuracy, hyperparameter optimization is used, and explainability is added with the help of permutation importance and Local Interpretable Model-agnostic Explanations (LIME), which allows finding the essential influencing factors, including closing, high, and low prices.

The proposed model, which is called LIM-GRLST, combines neural hybridization and weighted ensembling to address both long-term and short-term sequential trends of cryptocurrency time series. It is necessary to explain that this work does not bring a new recurrent mechanism, however, its value is the development of an integrated and explainable forecasting model that integrates hybrid deep learning with post-hoc explainability methods.

Other than predictive performance, the framework has a prospective and possible applicability in financial decision-making in highly volatile markets. Better forecasting accuracy can be used to facilitate trend awareness and aid in informed decision making in the face of uncertainty, and is particularly useful in trading and risk management. This study however, is interested in predictive modeling and does not directly deploy trading strategies or economic simulations; hence these implications are conceptually discussed as opposed to being empirically validated.

Regardless of these contributions, it is still not a guarantee that model generalization will be effective across time and different assets because cryptocurrency markets are very dynamics and non-stationary. This weakness is recognized and indicates the necessity of further research with several assets, larger datasets, and more powerful validation plans.

LIM in this work means a Local Interpretable Model, which is added to give instance-level explanations of the model predictions, contributing to its transparency and practical applicability in financial forecasting.

1.1 Research Contributions

The key contributions of this study are as follows:

i. A hybrid LSTM–GRU framework is developed to capture complementary temporal dependencies in cryptocurrency time-series data, and its performance is compared against multiple baseline models.
ii. Deep learning models (including LSTM, GRU, and hybrid variants) are systematically evaluated using standard performance metrics, namely MAE, MSE, RMSE, and R^2, to ensure comprehensive assessment.
iii. Statistical validation using the Diebold–Mariano (DM) testis conducted to verify the significance of performance differences between competing models.
iv. Model interpretability is enhanced through the application of permutation feature importance and LIME, providing feature-level insights into key predictors such as closing, high, and low prices.

v. The study presents a unified framework integrating prediction accuracy, explainability, and statistical validation, offering a more reliable and practical decision-support tool for cryptocurrency forecasting.

1.2 Research Hypotheses

Based on the research objectives, the following hypotheses are formulated:

- **H1**: The proposed hybrid LSTM–GRU model achieves significantly better forecasting accuracy compared to individual LSTM, GRU, and traditional hybrid models..
- **H2:** The performance improvements of the proposed hybrid model over baseline models are statistically significant.
- **H3:** Combining LSTM and GRU in a hybrid architecture captures complementary temporal features more effectively than standalone models.
- **H4:**The integration of explainability techniques provides meaningful insights into model predictions without degrading predictive performance.

Applying interpretability methods (permutation importance, LIME) makes it easier to understand behavior of the model, in the sense that none of the forecasting accuracy has been lost as a result of using interpretability methods.

2 Literature Review

The second generation of research after the initial Bitcoin-forecasting studies further expanded the modalities of data as well as the types of model, and the introduction of full surveys into the literature that presents a structured overview of the state of the art. The analysis of the relevant literature on the Bitcoin price prediction is presented in Table 1.

The study (Hamayel & Owda, 2021) factored socio-economic conditions with respect to Bitcoin price action and described a relative study using different parameters that influence forecasting. The outcome of DL models in the study means that explanatory variables other than pure price/volume need to be addressed explicitly in the design of cryptocurrency forecast. The authors in Aggarwal et al. (2019) reviewed cryptocurrency price-prediction algorithms and set out future directions. The main point they raise is that due to the volatility of crypto markets, it is difficult to predict the future development, although the toolkit of potential predictions increases in size with the introduction of newer technologies. The work standardizes the field and inspires common practices of evaluation. The study (John et al., 2024) was dedicated to short-term (hour-ahead) forecasting of Bitcoin with input variables that contain a Bitcoin price in USD, weighted price, and date. The formulation makes clear the sensitivity to near-term inputs and indicates how elementary market characteristics (price in USD and a weighted aggregation) can be used as parameters of the model at the intraday horizon.

Table 1 Overview and comparison of recent studies on Bitcoin price prediction

Ref. NO	Study / Model	Model type	Target / Forecast horizon	Key features	Findings
Farooq et al. (2024)	Farooq et al.	ADE TFT	Bitcoin (multi-horizon)	Attention + TFT for long dependencies	Lower MAPE, MSE, RMSE
Lee (2025)	Lee	TFT-based	Multi-crypto	On-chain + technical indicators	Better than classical baselines
Dincer and Kilimci (2025)	Dincer & Kilimci	Transformer vs LSTM	Multivariate	LSTM, TFT, GPT-2, Autoformer comparison	Transformers outperform others
Hasanli and Dursun (2026)	Hasanli & Dursun	Hybrid Transformer	Bitcoin (multi-step)	Transformer + LSTM/GRU	Hybrid outperforms recurrent
Peik et al. (2025)	Peik et al.	Adaptive TFT	High-frequency crypto	Dynamic segmentation	Better than fixed TFT/LSTM
Ud Din and Khan (2026)	Ud Din & Khan	TFT Hybrid	BTC closing price	TFT + BiLSTM + XGBoost	Strong in volatile markets

The study (Jiang, 2020) provides a review that contrasts statistical and ML approaches on recent approaches on cryptocurrency price prediction. The results of the analysis further reaffirm the importance of cross-family benchmarking, since the almost-irrelevant incremental improvement ML-based models have over classical approaches to Bitcoin represent more of a practical consideration than a strictly mathematical one. Kervancı and Akay (2020) proposed a DL approach to predict the price of Bitcoin where prices were quoted in the US dollar and data retrieved through Bitcoin price index. The empirical application points to the fact that DL models can be trained to directly forecast price-levels in customary currency specification, making the research practice to be similar to its application in the real world.

The authors presented a DL integration process on Bitcoin prediction, which approximates the external factors into the simplification of the prediction problem (Rizwan et al., 2019). They argue that by selecting apriori information relevant exogenous variables, the tractability of the model training can be improved. The study (Zhang et al., 2022) explored DL to predict cryptocurrency and claimed that identification of a good evaluation metric is very difficult and suggested new/different/alternative evaluation measures. This puts choice of metrics, and not model architecture, at the heart of credible comparisons. Pintelas et al. (2020) trained a DL model that includes on-chain data with daily BTC price. They incentivize on-chain variables by

observing direct correlations to BTC that can be useful in prediction and they place their approach within the larger context of multi-type data (prices plus blockchain signals).

The study (Kim et al., 2022) suggested a learnable window size with multilayer perceptron (MLP) to predict Bitcoin price. Rather than pre-determining the look-back window, the model learns the length of the window, since the most favorable market conditions may vary over time. Similarly, Rajabi et al. (2022) introduced a DL model to Bitcoin price prediction and successfully cast the problem as an analogue to finance/stock-market-type forecasting tasks. The analysis supports the view already prevalent in the field that DL models are good options for noisy and complex financial data. Singh et al. (2021) presented an approach for predicting Bitcoin price, relying on the logistic regression and reinforced with an LSTM. This combination set up reappears in the literature as a theme of blending a simple, interpretable linear component with a nonlinear recurrent learner that serves to capture nonlinear structure. The researchers compared four approaches to the price prediction of Bitcoin, logistic regression, support vector machine (SVM), recurrent neural network (RNN), and ARIMA and reported the accuracy of the approaches (Patel et al., 2022). The design is useful as it provides a station against which claims of DL can be anchored to statistical and classical ML approaches.

The study (Andi, 2021) suggested a DL-based framework with inter-relationship with regard to cryptocurrency price prediction. They observe technical and sentimental considerations, integrated by several researchers as well as pointing to the utilization of Twitter as a method of predicting market actions which positions sentiment as an auxiliary information source in crypto prediction. Although previous literature has used CNN, LSTM, GRU and transformer architectures, with or without other types of simple hybrid models, few studies have employed these architectures with an explainability framework to predict cryptocurrencies. As opposed to the previous methods, the proposed LIM-GRLST model in the current study, unites the feature extraction, deep temporal modeling, and interpretability in a single framework. Such hybridization is guaranteed to guarantee predictive performance and explainability, an essential gap in the literature.

Transformer-based and hybrid attention models have been progressively investigated in recent bitcoin forecasting research to get around recurrent architectures' shortcomings in managing volatile financial time series and capturing long-range dependencies. As demonstrated by studies that apply deep learning-enhanced TFT models for Bitcoin forecasting with reduced error metrics like MAPE and RMSE (Farooq et al., 2024), one noteworthy direction is the application of the Temporal Fusion Transformer (TFT), which integrates attention mechanisms with interpretable temporal dynamics to improve prediction accuracy and interpretability. Beyond single-asset prediction, TFT has also been used to forecast multiple cryptocurrencies by combining on-chain metrics and technical indicators (e.g., SOPR, TVL, RSI, and MACD), showing improved performance compared to RNNs and traditional baselines while informing trading strategies (Lee, 2025).

The superiority of attention mechanisms in capturing long-range temporal patterns across multivariate data has been demonstrated by comparative studies that go beyond specific model families and assess transformer-based designs alongside

LSTM and GPT-style models (Dincer & Kilimci, 2025). For multi-step Bitcoin forecasts, researchers have also suggested hybrid transformer techniques that combine transformer blocks with recurrent units, demonstrating improved trend tracking and reduced error values in comparison to solo recurrent models across 7-, 14-, and 21-day horizons (Hasanli & Dursun, 2026). Additional developments include customized frameworks that integrate TFT with additional attention-based modules and meta-learners to improve predictive robustness (Peik et al., 2025) and adaptive TFT approaches that dynamically segment time series for responsive forecasting in highly non-stationary crypto markets (Ud Din & Khan, 2026).

Recent work has shown that transformer- and attention-based approaches to cryptocurrency forecasting have been studied more, such as the Temporal Fusion Transformer (TFT), which can effectively capture long-range dependencies. These methods are usually very computationally complex and interpretable, however. Even though a few studies have include explainability methods, they are usually used in isolation and do not incur an integration in hybrid structures or with stringent statistical validation.

Simultaneously, some studies have explored hybrid LSTMGRU networks to predict time-series. Although such studies have shown an increase in the predictive accuracy, in general they offer little information on the model interpretability. The learning behavior of complementary nature of hybrid architectures is therefore not adequately warranted.

Thus, although there have been recent developments, current literature focuses on either predictive performance with transformer-based models that have a poor interpretability or present explainability without proper statistical validation. These drawbacks emphasize the importance of a single framework that incorporates hybrid modeling, explainable AI, and statistical validation to provide interpretable price predictions and forecasts of cryptocurrencies as well as precise ones.

2.1 Research Gap

Even though recurrent and transformer-based designs have made great strides in bitcoin forecasting, there are still a number of unanswered questions. First, explainability techniques and hybrid architectures with attention or transformer models are not well integrated, leaving practical interpretability mainly unexplored. Second, the majority of research lacks thorough statistical validation over a range of forecast periods, which is essential for evaluating model dependability in erratic markets. Third, recurrent models and attention-based architectures have not been sufficiently compared for short- and multi-step prediction tasks. The current study aims to close these gaps by proposing a hyperparameter-optimized hybrid LSTM–GRU model that combines permutation significance and LIME to provide both prediction accuracy and comprehensible insights for bitcoin decision assistance.

3 Proposed Methodology

Figure 1 shows that the proposed study uses a layered hybrid DL approach, LIM-GRLST, which combines the LSTM and GRU networks to learn long-term and short-term temporal relationships in cryptocurrency time series. First, the raw data are processed, such as normalization, generation of sequences and reduction of noise. Parallel LSTM and GRU layers are fed with the preprocessed sequences and LSTM is effective in modeling long-time dependencies, while GRU is efficient in short-term dynamics and faster training. The results of these layers are then fused using a strategy of neural hybridization, which is also such that the most informative aspects of each network are maintained. Hyperparameter tuning is used to optimize the learning rate, the number of units, the batch size and the dropout rates. Moreover, the model interpretability is increased with the help of permutation importance and LIME, which reveal the importance of input features and confirm the accuracy of the predictions. Lastly, the proposed LIM-GRLST model is compared with baseline and existing models on metrics of performance like mean absolute error (MAE), root mean-square error (RMSE), and R^2, which reveal its strength and predictive power on forecasting Bitcoin price. The proposed model combines LSTM and GRU in a parallel structure to leverage their complementary learning capabilities, followed by an explainability layer to interpret model predictions. The LIM component is integrated to enhance interpretability by generating local explanations for individual predictions, enabling a better understanding of model behavior in complex financial time-series data.

The proposed LIM-GRLST model is the combination of both LSTM and GRU networks that formulates effective sequential dependencies of time-series cryptocurrency data. The model feeds on an input sequence of shape (30, 5) and this corresponds to 30 time steps, 5 features per step. The architecture is separated into two parallel branches:

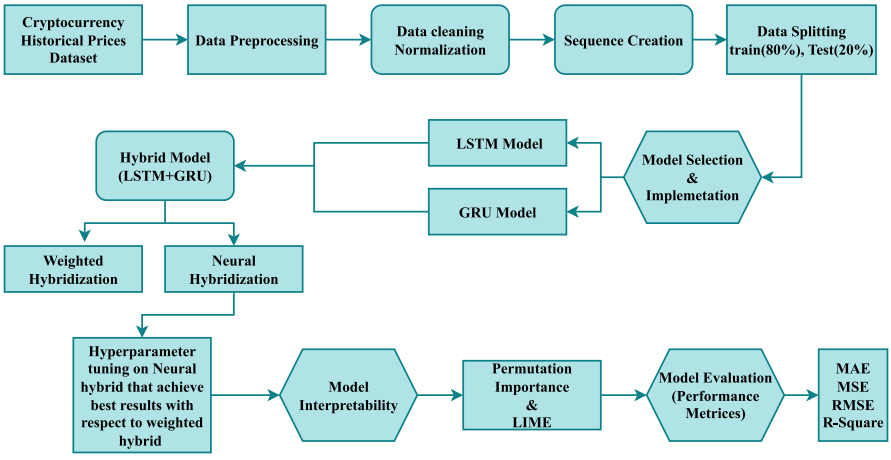


Fig. 1 Bitcoin price prediction pipeline implementing an interpretability-enhancing tuned neural hybrid LIM-GRLST model with LIME and permutation importance

- Descriptor LSTM branch: A 64-unit, 32-unit two-layer LSTM network which learns long-term temporal dependencies.
- GRU branch: The two-layer GRU network consisting of 64 and 32 units to capture shorter and more recent dependencies with a lower cost of computation.

The two branches give their outputs, which are fused in a fusion layer, which is a combination of the learn representations. This is then succeeded by a layer with 16 ReLU units in order to learn higher-level features and finally an output layer with a single unit that produces the predicted price. This hybrid architecture utilizes the complementary learning of LSTM and GRU that leads to a higher level of prediction accuracy than with either architecture).

3.1 Dataset Description

This study leverages the benchmark dataset entitled Cryptocurrency Price History accessible on what is known as Kaggle (Rajkumar, 2025). The dataset used in this study consists of daily Bitcoin prices from April 29, 2013, to July 6, 2021. Only the features relevant to the model, Open, High, Low, Close, and Volume—are included. Table 2 summarizes their descriptive statistics, providing a concise overview of the data without unnecessary detail. It should be noted that the analysis is limited to Bitcoin and does not include other cryptocurrencies, which constrains the generalizability of the findings to different market regimes or post-2021 periods. Future work may extend this framework to multiple cryptocurrencies and more recent data to evaluate broader applicability (Kaur et al., 2025b; Seabe et al., 2023).

3.2 Data Preprocessing

Preprocessing of data is one crucial procedure in making the dataset, reliable and suitable in training deep learning models. The steps below were carried out:

3.2.1 Handling Missing Values and Duplicates

The data were analyzed under the missing values. Rows with null values were eliminated so as to ensure that data integrity would not be compromised. The duplication of entries was also noted and was away with to avoid skewed analysis and training of models.

Table 2 Descriptive statistics of bitcoin daily prices (2013-2021)

Feature	Count	Mean	Std	Min	Max
Open	2991	6700.15	11288.04	68.50	63523.75
High	2991	6893.33	11642.83	74.56	64863.10
Low	2991	6486.01	10869.03	65.53	62208.96
Close	2991	6711.29	11298.14	68.43	63503.46
Volume	2991	1.09e10	1.89e10	0.00	3.51e11

3.2.2 Feature Selection and Transformation

The column of date was changed to a datetime and used as an index of time series analysis. Unneeded columns like Name, Symbol, and Marketcap were also discarded since they did not directly enter as input values into models that were implemented in this study. The categories of other numerical characteristics (High, Low, Open, Close, Volume) were taken to model.

3.2.3 Feature Scaling

The numerical features were normalized to enhance the stability of the training of the neural networks by using Min-Max Scaling. MinMaxScaler was applied exclusively to the training set before transforming the test set, preventing information leakage. In this approach, the value of the different features is scaled to the range of 0 to 1 as shown in Eq. 1:

$$X_{\text{scaled}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{1}$$

where X is the original feature value, $X_{\min}$ is the minimum value of the feature, and $X_{\max}$ is the maximum value of the feature.

3.2.4 Time Series Sequence Creation

Training of models like LSTM and GRU requires that the data be arranged in sequences. A function was provided to generate feedback sequences (X) and target values (y). The input sequences would consist of some scaled features of some set number of previous time points (seq_length) and output would be closing price at the following time step. The sequences were developed in the following ways: Given a time series dataset having N time steps and F features, and a sequence length L, the input sequences X are built so that each item X_i is a 3D array of shape (L, F) where i runs over [0, N-L-1]. The relevant target value y_i is the feature value (here, Close) at time step $i + L$ as Eq. 2.

$$X = [x_0, x_1, \ldots, x_{N-L-1}], \quad y = [y_L, y_{L+1}, \ldots, y_{N-1}] \tag{2}$$

where x_i is the sequence of features from time i to $i + L - 1$, and $y_i + L$ is the target value at time $i + L$.

3.3 Train-Test Split

The preprocessed data was divided into training and testing sets. After cleaning and scaling, the dataset contains 2991 daily records. Sequences of 30 past days were created for one-step-ahead prediction, resulting in 2961 sequences split into training and test sets. The division into training and testing sets was also done in the form of

a temporal split, the first 80 percent of the data was considered training and the 20 percent- remaining as testing data.

3.4 Learning Models

This section discusses how some deep learning models can be used to predict the prices of Bitcoin.

3.4.1 Long Short-Term Memory

LSTM networks are a form of RNN that are developed to deal with sequential information and long-range dependencies. An LSTM cell is composed of a number of gates that regulate the flow of information: the input gate, the forget gate, the output gate. These gates enable the LSTM to selectively recall or exclude past material and therefore it is good at time series forecasting. The structure of LSTM model used is:

- An LSTM layer with 64 units, returning sequences.
- An LSTM layer with 32 units.
- A Dense layer with 16 units and ReLU activation.
- An output Dense layer with 1 unit for the price prediction.

The equations of LSTM layer are complicated in nature, yet the fundamental concept is focused at cell state (C_t) and the hidden state (h_t).

Forget Gate Determines what information to throw away from the cell state as in Eq. 3.

$$ft = \sigma\left(Wf \cdot [ht-1, xt] + bf\right) \tag{3}$$

Input Gate Decides which new information to store in the cell state as in Eq. 4.

$$\begin{aligned} it &= \sigma\left(Wi \cdot [ht-1, xt] + bi\right) \\ \tilde{C}t &= \tanh\left(WC \cdot [ht-1, xt] + bC\right) \end{aligned} \tag{4}$$

Update Cell State Updates the cell state as Eq. 5.

$$Ct = ft \odot Ct - 1 + it \odot \tilde{C}t \tag{5}$$

Output Gate Determines the output based on the cell state in Eq. 6.

$$ot = \sigma\left(Wo \cdot [ht-1, xt] + bo\right), \quad ht = ot \odot \tanh(Ct) \tag{6}$$

where σ is the sigmoid function, $\odot$ is element-wise multiplication, W and b are weights and biases, h_{t-1} is the previous hidden state, and x_t is the current input.

3.4.2 Gated Recurrent Unit

Another RNN model is GRU, which is slightly simpler than LSTM, having fewer gates. It merges the forget and the input gates into an update gate and further has a reset gate as well. The RUs can be seen to be roughly equivalent to LSTMs in terms of performance however computationally cheaper. The architecture of used GRU model includes:

- A GRU layer with 64 units, returning sequences.
- A GRU layer with 32 units.
- A Dense layer with 16 units and ReLU activation.
- An output Dense layer with 1 unit for the price prediction.

The key equations for a GRU are:

Update Gate Determines how much of the past information to pass to the future as in Eq. 7.

$$zt = \sigma\left(Wz \cdot [ht-1, xt] + bz\right) \tag{7}$$

Reset Gate Determines how much of the past information to ignore as in Eq. 8.

$$r_t = \sigma\left(Wr \cdot [ht-1, xt] + br\right) \tag{8}$$

Candidate Hidden State Computes a new candidate hidden state as in Eq. 9.

$$\tilde{h}t = \tanh\left(Wh \cdot [rt \odot ht-1, xt] + bh\right) \tag{9}$$

Update Hidden State Updates the hidden state as in Eq. 10.

$$h_t = (1 - zt) \odot ht - 1 + zt \odot \tilde{h}t \tag{10}$$

Similar to LSTMs, σ is the sigmoid function, tanh is the hyperbolic tangent function, $\odot$ is element-wise multiplication, W and b are weight matrices and bias vectors, h_{t-1} is the previous hidden state, and x_t is the current input.

3.5 Proposed Hybrid Models

Hybrid models are a combination of various models to take advantage of the strengths of each. The term hybrid may also apply to neural architectures and systems where

two or more different types of neural networks, such as GRU and LSTM, can be used with each other, and/or deep learning techniques may be applied in combination with other statistical methods. The aim is to typically find other patterns in data that may not have been detected by a single model and thereby achieve a better performance. The paper examines two kinds of hybrid models, as illustrated in Fig. 2

3.5.1 Hybrid Weighted Average Model

A hybrid model was developed by averaging the results of the majority of the weighted individual GRU and LSTM models. The weights were experimentally selected (0.7 on GRU and 0.3 on LSTM) depending on their respective performances according to Eq. 11.

$$\hat{y}_{hybrid_weighted} = wGRU \cdot \hat{y}GRU + wLSTM \cdot \hat{y}LSTM \tag{11}$$

where $\hat{y}GRU$ and $\hat{y}LSTM$ are the predictions from the GRU and LSTM models, respectively, and $wGRU$ and $wLSTM$ are their corresponding weights ($wGRU + wLSTM = 1$).

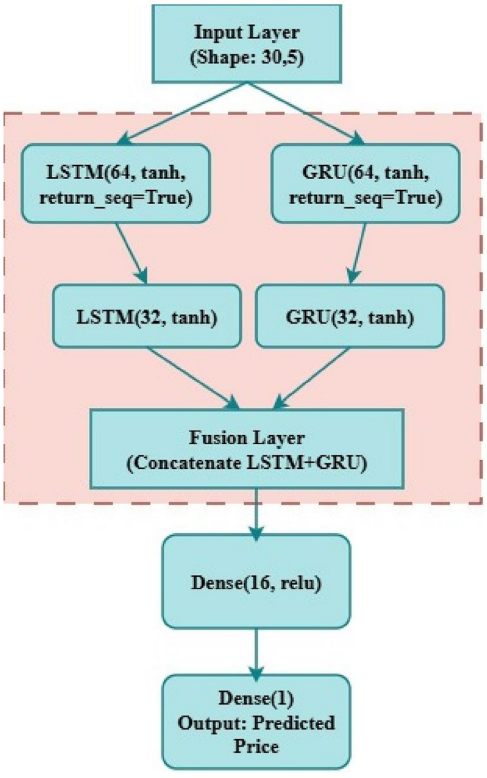


Fig. 2 Layered architecture of the proposed LIM-GRLST hybrid model. The model consists of parallel LSTM and GRU branches that process sequential input data to capture both long-term and short-term temporal dependencies. The outputs of both branches are concatenated in a fusion layer, followed by dense layers for final prediction of Bitcoin price. This hybrid structure enhances learning capability by combining complementary recurrent representations

3.5.2 Neural Hybrid (LIM-GRLST Concatenate) Model

A more complex neural hybrid model was the one created with the combination of the GRU and LSTM layers into one block. Through this model, the network can pick on what is happening between the features in the two branches composed of GRU and LSTM. The structural design of the Neural Hybrid model is as illustrated below:

- A shared Input layer taking the time series sequence.
- A GRU branch with two GRU layers.
- An LSTM branch with two LSTM layers.
- A Concatenate layer merging the outputs of the final GRU and LSTM layers.
- Dense layers with ReLU activation (including hyperparameter tuning for the number of layers and units) applied to the concatenated output.
- An output Dense layer with 1 unit.

This model sees how to combine the representation learned by GRU and LSTM networks and in this way, it is likely to capture even broader patterns in the time. Hyperparameter tuning, such as the specific amount of units and layers in the dense part, the activation function, the learning rate and the dropout rate, was performed. Hybrid models are a combination of various models to take advantage of the strengths of each. The term hybrid may also apply to neural architectures and systems where two or more different types of neural networks, such as GRU and LSTM, can be used with each other, and/or deep learning techniques may be applied in combination with other statistical methods. The neural hybrid (LIM-GRLST) architecture combines LSTM for long-term dependencies and GRU for computational efficiency. This hybrid is tuned specifically for cryptocurrency forecasting, and the descriptive label emphasizes the practical combination rather than a novel theoretical contribution. The aim is to typically find other patterns in data that may not have been detected by a single model and thereby achieve a better performance. The paper examines two kinds of hybrid models, as illustrated in Fig. 2.

3.6 Hyperparameter Tuning

The Neural Hybrid model's performance was further optimized through hyperparameter tuning using the Keras Tuner library. A RandomSearch tuner was employed to explore a defined search space for hyperparameters including the number of units in GRU and LSTM layers, the number and size of dense layers, activation function, learning rate, and dropout rate. The tuner was configured to minimize the validation loss. During hyperparameter tuning, model performance was monitored on a held-out portion of the data to guide the optimization process. To ensure proper generalization and avoid data leakage, all models were trained using chronologically split datasets. Early stopping and dropout regularization were applied, and sequences were created to predict only subsequent targets, preserving temporal order. The best hyperparameters found were then used to train the final neural hybrid model as shown in Table 3.

Table 3 Hyperparameter tuning parameters for the neural hybrid model

Parameter	Type	Search space / Values	Description
`gru_units`	Integer	[32, 64, 96] (step=32)	Number of units in the first GRU layer
`lstm_units`	Integer	[32, 64, 96] (step=32)	Number of units in the first LSTM layer
`dense_layers`	Integer	[1, 2]	Number of dense layers after concatenation
`dense_units`	Integer	[16, 32, 64] (step=16)	Number of units in the dense layer(s)
`activation`	Choice	{relu, tanh}	Activation function for dense layer(s)
`learning_rate`	Choice	[0.001, 0.005, 0.01]	Learning rate for the Adam optimizer
`dropout_rate`	Float	[0.0, 0.1, 0.2] (step=0.1)	Dropout rate after dense layer(s)
`batch_size`	Choice	[32, 64]	Batch size for training
`epochs`	Choice	[30, 50]	Maximum number of epochs for each tuner trial

3.7 Model Interpretability

Interpretability of a model is the capacity to comprehend how exactly a model makes a prediction. Interpretability methods are important tools of trust, debugging of models, and scientific progress, especially with complex black-box models, such as deep neural networks. This study uses two methods for the interpretation.

i. **Permutation Importance:** The relevance of a feature is then gauged according to how much its values in the validation or test data limit model performance. Shuffling a feature and making the performance of the model deteriorate greatly, is a sign that it was an important feature. Permutation importance offers a global features importance across the whole dataset.
ii. **LIME:** PIM can also be written as a technique that describes guesses of any black-box model by summarizing them through fitting a local, understandable model at the prediction point. On a particular instance, LIME alters the instance and uses the black-box model to predict the altered instance; afterwards, a weighted training is performed with the altered instance and predictions based on the closeness to the initial instance. The local interpretable model has weights that identify how various features lend significance to that particular forecast. In traditional form,

LIME is locally interpretable, that is, what the model predicts about an individual instance, not the overall model.

3.8 Performance Evaluation Metrics

A number of statistical metrics of performance are used to evaluate the efficiency of the proposed cryptocurrency price prediction framework. These statistics give analytical data to the precision and stability of the forecasting model. The MAE, MSE, RMSE, and R^2 inside metrics are used regularly in predicting time series data.

i. **Mean Absolute Error:** The Mae provides measurement of the average magnitude of errors with no regard to their direction. It provides direct interpretation regarding the extent of average nature of the deviation of the predictions compared to the actual values. A lower MAE implies higher accuracy of prediction.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{12}$$

In Eq. 12, *n* denotes the total number of observations, y_i is the present value of Bitcoin. To get hold of online programs, it is important that both the underestimation and close estimation affect the error equally. The sum summates these errors to all the measurements and n is then divided thus, giving the average deviation. MAE is also written in the same units as the target variable, so there is a simple interpretation of prediction error. The lower MAE, the better the prediction accuracy, which means that the forecasts of the model will be more likely close to reality pictures.

ii. **Mean Squared Error:** The average of the squared differences between the predicted results and the actual results is obtained by MSE. The square of the errors means that larger errors carry a greater penalty and it is especially good at detecting large prediction errors.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{13}$$

In Eq. 13, *n* denotes the number of all the observations, y_i is the real price of the Bitcoin and $\hat{y}_i$ is the expected one. The error of each data point $(y_i - \hat{y}_i)$ is squared because magnitude does not matter; negative as well as positive contributions are summed. The summation of squared errors is done on n number of observations and the MSE value is calculated as a mean of the summation. MSE is very sensitive to anomalies affecting the score because in squaring operation, larger differences contribute in a greater extent to the final score. The smaller the MSE the better the model the smaller the difference between the predicted value and the actual value.

iii. **Root Mean Squared Error:** MSE can be simplified as RMSE which is the most frequently used error measure in forecasting. It is convenient because the errors will be expressed in the same unit of measure as the data, therefore, easier to interpret. RMSE is lower which shows greater accuracy of the model according to Eq. 14.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{14}$$

iv. **Coefficient of Determination** (R^2): R^2 is used to indicate the amount of variance in the dependent variable that is estimated by the model. It goes between 0 and 1 with higher scores closer to either 0 or 1 signifying a greater degree to which the model explains more of the variance in the data.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{15}$$

In Eq. 15, y_i is the envisaged Bitcoin price, $\hat{y}_i$ is the expected and y is the mean of all y_i. The formula will be the mean squared error (MSE), which is represented in the numerator whereas the denominator will measure the variance in the actual prices. The R^2 is between 0 and 1. A model with a high R^2 (near 1) is tell-tale when the price is replicated however a low R^2 indicates weak forecasting power of the model. This is a very important measure of assessing goodness-of-fit and the accuracy of a model in predicting trends in the price of Bitcoin.

3.9 Experimental Setup

This subsection explains the experimental setup that was used in this research. Every experiment was implemented in Python (version 3.x) in the Google Colaboratory platform that offers an online environment to execute and test deep learning models in the cloud. The packages used included TensorFlow (v2.19.0) and Keras (v3.10.0) to develop the models.

It was supported by NumPy (v2.0.2), Pandas (v2.2.2) to process data, Scikit-learn (v1.6.1) to preprocess and evaluate, Matplotlib and Seaborn to visualize, Keras-Tuner (v1.4.7) to optimize hyperparameters and LIME (v0.2.0.1) to interpret a model. These are used to guarantee the reproducibility and transparency of the experimental pipeline.

Random seeds were fixed explicitly in NumPy, TensorFlow and the built-in Python random module to guarantee reproducibility. Nevertheless, small fluctuations are still possible since deep learning optimization has its own stochastic processes.

To maintain the temporal dependencies in the data, it was divided sequentially; 80 percent of the data was used to train and 20 percent to test. There was no shuffling to make sure that no future information was used in training the model.

The proposed Neural Hybrid model was trained and tested with five independent runs to measure the statistical strength. Mean and standard deviation are used to report performance measures to represent variability and guarantee valid assessment.

The standard regression measures, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and coefficient of determination (R2) were used to measure model performance in a comprehensive manner to evaluate predictions accuracy. While a fixed train-test split was used for evaluation, we acknowledge that more advanced time-series validation techniques, such as rolling window and walk-forward validation, can further enhance robustness. This has been clearly admitted and will be tackled in upcoming work.

4 Results and Discussion

This section gives the performance analysis of the proposed LIM-GRLST hybrid model to predict the price of Bitcoin. The models analyzed are standalone LSTM, standalone GRU, weighted hybrid, neural hybrid and the last tuned hybrid. Standard regression measures are used to evaluate all models such as Mean Absolute error (MAE), Mean Squared error (MSE), Root mean squared error (RMSE) and coefficient of determination (R2) to facilitate a holistic evaluation of the accuracy of the predictions and variations in errors.

It is evaluated with data on Bitcoin price between 2013-2021, and a chronological traintest split (80,20), which produces 598 test sequences. This makes sure that the model is tested using purely unobserved future data. All preprocessing operations, such as normalization are done in such a way that they do not leak information.

It has a high predictive performance with RMSE = 0.0279, MAE = 0.0209 and $R^2 = 0.9897$ in a one-run experiment. Although these findings are encouraging in terms of good goodness-of-fit, the large value of the (R2) should be interpreted with care since the price-level data of cryptocurrencies is highly autocorrelated and persistent.

Five independent runs were conducted to make the model robust and to consider the issues of overfitting or random starting. The averaged performance is reported as: Mean $R^2 = 0.9236 \pm 0.0165$, MSE = 0.0626 ± 0.0135, and MAE = 0.1359 ± 0.0153 . These outcomes give a more realistic estimation of model performance and show that it is stable in terms of run-to-run.

In order to further confirm the efficacy of the proposed model, they were compared against baseline methods. The Naïve (Random Walk) model achieved $R^2 = -1.2037$, MAE = 0.9931, and MSE = 1.8057, while the ARIMA model resulted in $R^2 = -0.6227$, MAE = 0.7355, and MSE = 1.3297. The much worse performance of these baselines supports the fact that the suggested hybrid model removes meaningful nonlinear temporal patterns other than mere persistence.

The performance improvement is statistically validated by the Diebold-Mariano (DM) test in which all the comparisons (Hybrid vs LSTM, GRU, and Weighted Hybrid) showed that all the values were below the significance threshold of $p < 0.001$.

In terms of modeling, the complementary benefits of the LSTM and GRU structures in the creation of both long-term and short-term sequential structures explain

the strong performance. Hyperparameter optimization also improves the predictive power and generalization of models.

Besides predictive performance, the findings can also be studied to give valuable insights into the behavior of the cryptocurrency market. The fact that the features used to include closing price, high, and low values dominate suggests that the model is able to capture short-term momentum and volatility, which are central to financial time-series dynamics. The incorporation of interpretability techniques also enhances feature impact knowledge, which can help in making better decisions in fields like trading, portfolio optimization, and risk management.

Nevertheless, there are a few limitations that need to be considered. The analysis is limited to a particular timeframe (2013–2021) and a single asset (Bitcoin), which can be a limitation to generalizing the results to other market regimes. Cryptocurrency market is a non-stationary structure that is susceptible to structural shifts, which are a result of external forces like regulations, macroeconomic trends, and investors.

In addition, the present analysis is on a fixed traintest split. More sophisticated validation plans (rolling window or walk-forward validation) are required to further test model robustness amidst changing market conditions. Exogenous variables can also be used in future work, e.g. macroeconomic variables and sentiment data can be integrated to capture more complex market behavior.

In general, the suggested LIM-GRLST model has a high predictive power, which is proved by the robustness analysis, comparison with the baseline, and statistical validation, but the limitations and use in practice of any financial forecasting are transparent.

4.1 LSTM Model Results

The LSTM was used as a baseline since it has been proven to be effective in modeling long-range dependencies in sequential data. The LSTM recorded a value of 0.0445, 0.0055, 0.0741 and 0.9277, MAE, MSE, RMSE and R^2 respectively. Results show that the model was capable of learning meaningful time-dependent representations albeit the error being rather high when compared to that of other models which implies the presence of certain limitations in handling complex time-varying variations. The LSTM was prone to slight overfitting based on the deviating loss on the validation sets in various epochs. This confirms that in as much as LSTM can be effective, given that it is complex and requires extensive parameter sets in its implementation, its usefulness can be curbed in the case of some forecasting applications as illustrated in Fig. 3.

4.2 GRU Model Results

The GRU performed better in most of the measurements with an MAE of 0.0381, MSE of 0.0048, RMSE of 0.0690 and R^2 of 0.9373. In comparison to LSTM, GRU also has fewer gates, so the computational cost is lower, nevertheless with the ability to capture long-term dependencies. This effectiveness was equated with quicker convergence and improved generalization, which is denoted through the reduced error scores and a high value of R^2. The success compared to LSTM shows that the simpli-

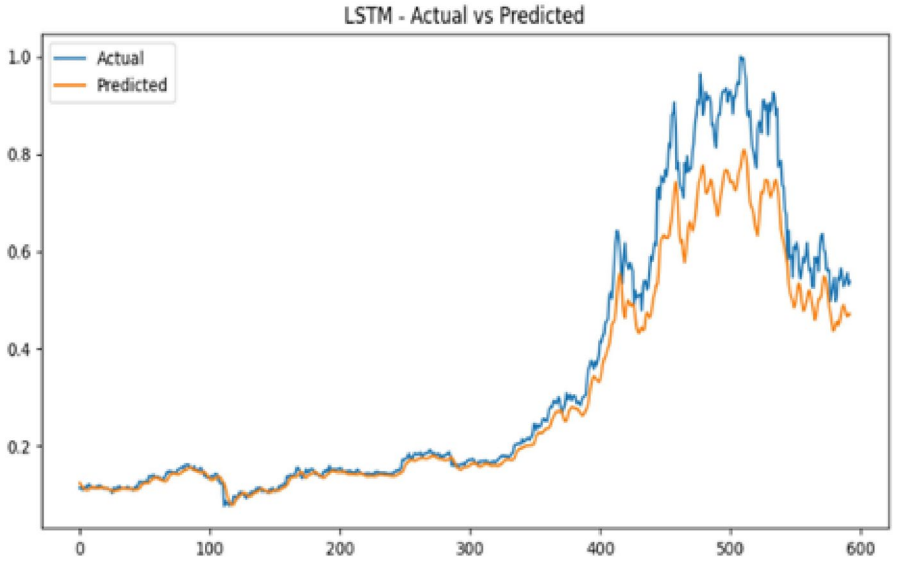


Fig. 3 Actual vs. predicted Bitcoin price trends using the LSTM model. While the model successfully captures the overall upward trend and major price movements, noticeable deviations are observed during periods of high volatility, indicating limited capability in modeling sudden fluctuations

fied gating mechanism of GRU has been proven successful in the financial and time series applications, where speed and accuracy are paramount factors. Although, GRU was more accurate than LSTM, its accuracy reached a plateau before LSTM, therefore, additional improvements might need further integration as exhibited in Fig. 4.

4.3 Hybrid Model (LIM-GRLST, Weighted Fusion)

In an attempt to investigate what is best about LSTM and GRU, a weighted hybrid model was built, with a weighted average of their predictions. MAE of the hybrid is 0.0400, MSE is 0.0049, RMSE is 0.0703 and $R^2 = 0.9348$. Though this model offers a slightly better stability than standalone LSTM, it did not outperform that of the GRU. This shows that mere output-level fusion does not exploit the possible synergies between the two recurrent architectures comprehensively. The findings indicate that weighted fusion is able to smooth the predictions however cannot be used to extract deeper joint representations, therefore, constrained in its performance. This result triggered the inspiration to design a neural hybrid solution, which captures feature level interaction rather than just at the prediction level that is depicted in Fig. 5.

4.4 Neural Hybrid Model (LIM-GRLST Concatenation)

The neural hybrid model, which combined the hidden feature representations of both LSTM and GRU combining them in a concatenation prior to feeding to dense layers, provided a great performance gain. It obtained MAE 0.0285, MSE 0.0027, RMSE 0.0520, and R^2 0.9644. By integrating the properties of long-term dependency in

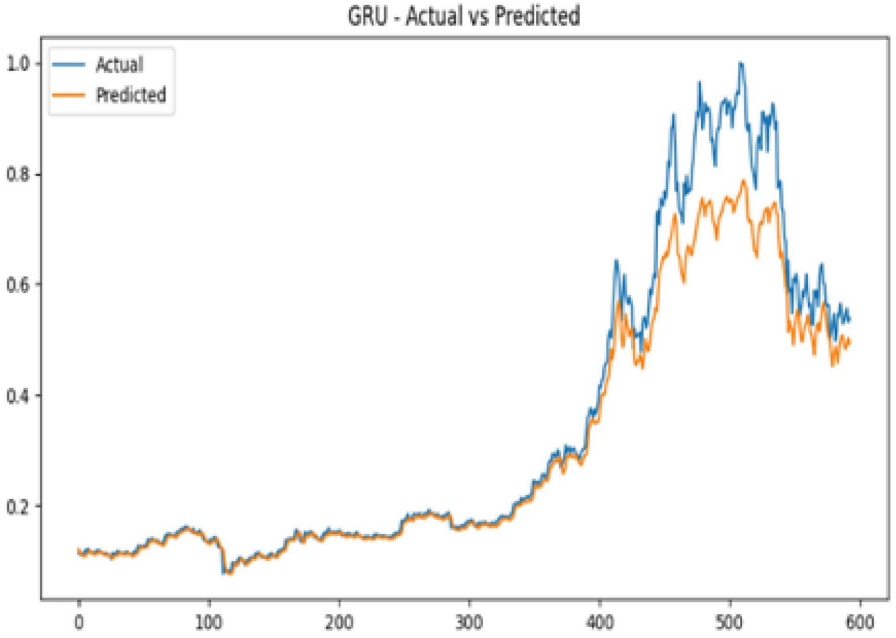


Fig. 4 Actual vs. predicted Bitcoin price trends using the GRU model. Compared to LSTM, GRU demonstrates slightly improved alignment with actual values and better generalization, though minor lag and smoothing effects are still visible during rapid market changes

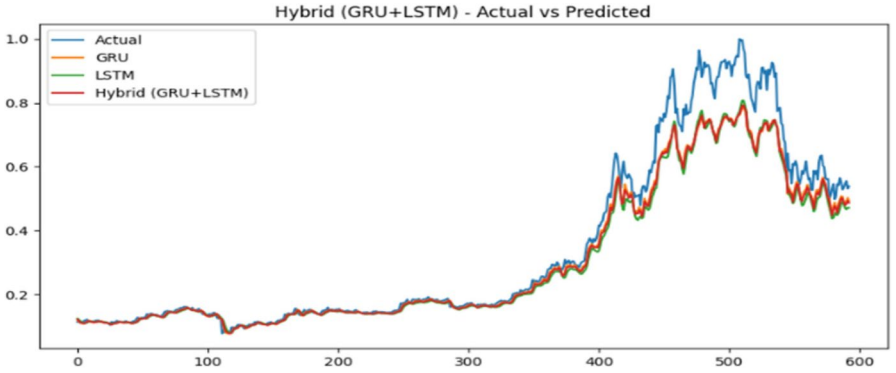


Fig. 5 Price predicted graph of weighted fusion (LIM-GRLST) for Bitcoin

LSTM and the efficiency of GRU this hybrid model has been shown to perform better in capturing the longer-term effects and short-term variations. The difference between the weighted and feature-level integration was that the latter allowed a richer representation and enhanced the generalization capabilities on new data. The improvement in error measures indicates the benefit to developing hybrid systems trading at the representational level as opposed to trading at the prediction level, which is consistent with current trends in DL-based time series forecasting as illustrated in Fig. 6.

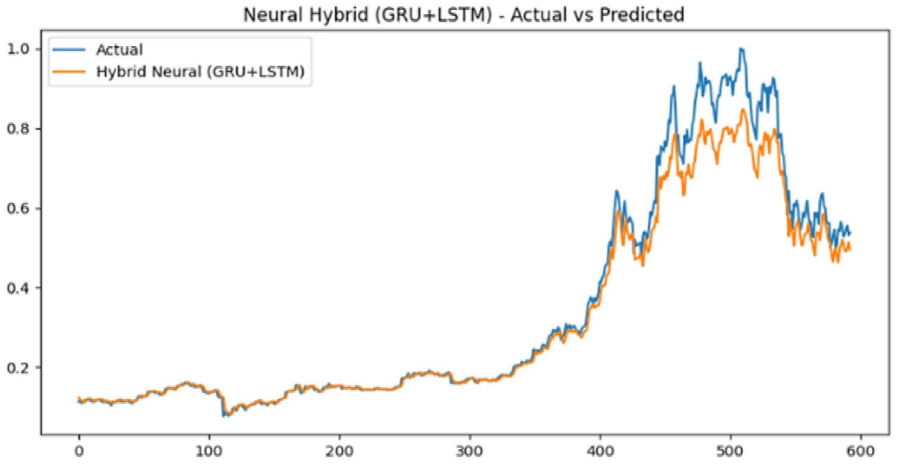


Fig. 6 Actual vs. predicted Bitcoin price trends using the proposed hybrid model. The model exhibits closer tracking of real price movements, especially during volatile regions, highlighting its enhanced capability to capture both short-term fluctuations and long-term dependencies

4.5 Tuned Neural Hybrid Model (Bayesian Optimization)

Lastly, the model in the form of a neural hybrid was optimized in terms of hyperparameters using the Bayesian search which returned the best setting (GRU units: 32, LSTM units: 32, Dense layers: 2, Dense units: 48, activation: tanh, learning rate: 0.001, drop out: 0.1). The tuned model obtained the MAE of 0.0209, MSE of 0.0008, RMSE of 0.0279, and R^2 of 0.9897, which means that the model showed a significant increase in accuracy relative to all other models. The close-to-zero R^2 value shows that the tuned hybrid model was able to explain nearly all the variance in target data. The augmentation brings out the importance of hyperparameter optimization in improving neural architectures of prediction. Moreover, the results of the performance shows that tuned hybrid models reduce prediction error and generalize very well thus highly applicable to real world financial forecasting applications as depicted in Fig. 7.

4.6 Sequence Length Sensitivity Analysis

A sensitivity analysis was performed with sequence lengths of 15, 30, 45, and 60 days and results are given in Table 4. The results indicate that a sequence length of 30 provides optimal performance, yielding the lowest MAE, MSE, RMSE, and the highest R^2. This provides a quantitative justification for the chosen sequence length in the neural hybrid model.

4.7 Diebold–Mariano test

The Diebold-Mariano test was performed between the proposed hybrid model and the baseline models to determine whether the changes in performance are statisti-

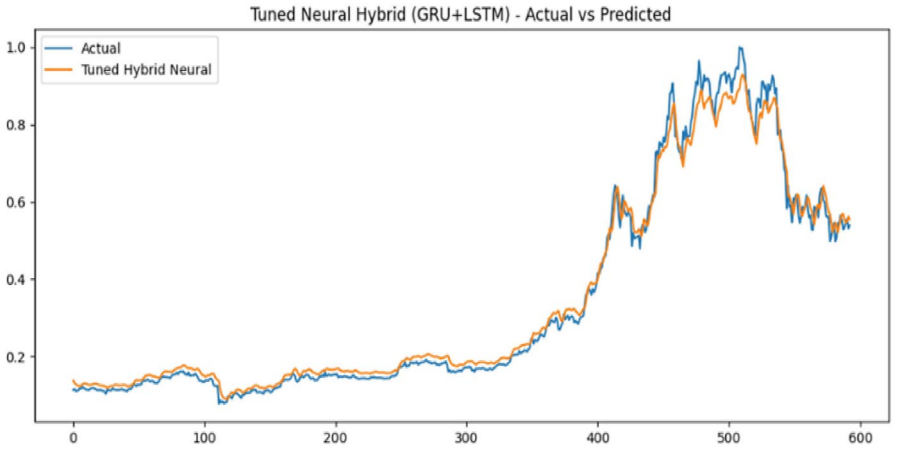


Fig. 7 Actual vs. predicted Bitcoin price trends using the tuned hybrid model with optimized hyperparameters. The predictions show minimal deviation from actual values, demonstrating improved stability and accuracy, particularly in capturing sharp peaks and drops in the market

Table 4 Sequence length sensitivity analysis for neural hybrid model

Sequence length	MAE	MSE	RMSE	R^2
15	0.0385	0.00350	0.0592	0.9576
30	0.0269	0.00163	0.0403	0.9803
45	0.0336	0.00288	0.0536	0.9652
60	0.0305	0.00227	0.0477	0.9726

Table 5 Diebold–Mariano test results

Comparison	DM statistic	p-value
Hybrid vs LSTM	−11.68	0.000
Hybrid vs GRU	−10.15	0.000
Hybrid vs Weighted Hybrid	−11.01	0.000

cally significant. The findings, as shown in Table 5, also indicate a strongly negative DM statistics with a p-value of less than 0.05, which shows that the proposed model highly performs in comparison with LSTM, GRU, and weighted hybrid models. The statistical significance test also reinforces the validity of the model proposed, that the higher performance is not accidental or due to the random changes in the data.

4.8 Permutation Importance Analysis

Following the individual input feature contributions, permutation importance was obtained by noting the change in MSE that occurred after each individual feature was shuffled. The results demonstrated that, close price had the most influential predictions which created an addition of 0.0244 in MSE when it was permuted. The Low (0.0092), High (0.0070) prices had the greatest influence with Open (0.0010) and Volume (0.00001) comparatively lesser impact. The results are logical since the closing price is generally the most vigilant feature of the financial markets that rep-

resents the investor sentiment and trading momentum, thus as a valuable short-range prediction factor. Permutation importance identifies the most influential predictors: closing, high, and low prices. Systematic feature ablation is planned for future work to further assess feature contributions. The slightly reduced significance of trading volume as an explanatory variable hints at the fact that cryptocurrency markets are price-dominated at least in the given data set as depicted in Fig. 8.

4.9 LIME Analysis

LIME were used on the final tuned hybrid model to give instance-level interpretability. Its output was the identification of the most important characteristics to the prophecy of particular test samples, with results shown in Fig. 9.

- High_t-0, Close_t-0 and Low_t-0 were the most influential features in the first test instance (True = 0.1139, Predicted = 0.1148), with minor negative impact of High_t-4 (-0.0051) and Close_t-3 (-0.0048).
- In another example (True = 0.1256, Predicted = 0.1160), the influence of recent close, high, and low was large, whereas the lags of older variables were quite small and negative.
- The same pattern was observed in the third case (True = 0.1339, Predicted = 0.1262) wherein Close_t-0 (0.0171) and High_t-0 (0.0161) were the most affective.

These findings clearly indicate that the model will mostly use the latest closing, high, and low prices to generate accurate forecasts, however the older lags and volume trading location have insignificant impacts. This is congruent with the understanding that market conditions in the recent past influence short-term price variability of crypto-currencies, and thus are employable materials in forecasting models.

The two main insights draw out of the interpretability results are: Excluding leverage, the permutation importance analysis shows that price related features (particu-

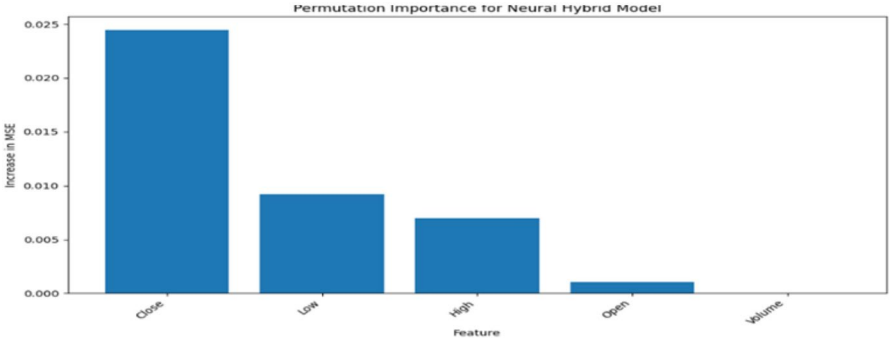
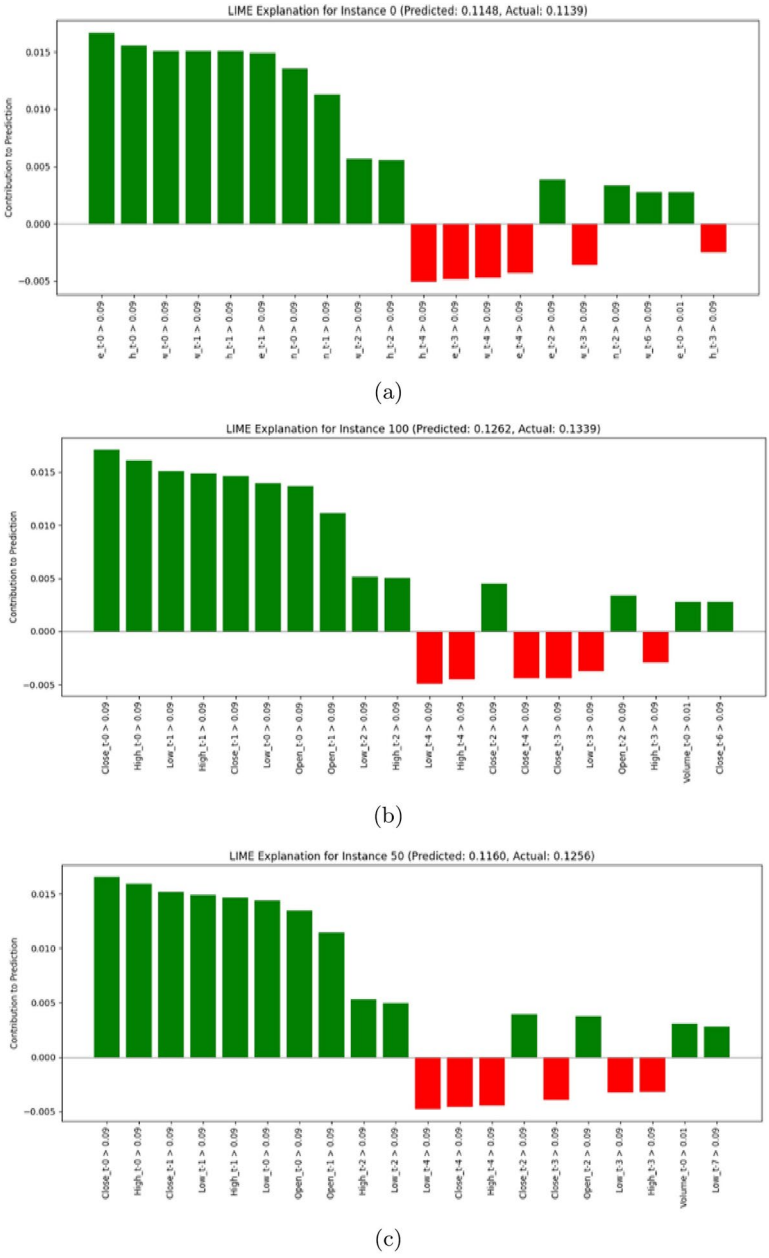


Fig. 8 Feature importance analysis using permutation importance for Bitcoin price prediction. The importance of each feature is measured by the increase in prediction error when the feature values are randomly shuffled. Results indicate that the closing price has the highest influence on model predictions, followed by low and high prices, while other features contribute comparatively less to the forecasting performance



(a)

(b)

(c)

Fig. 9 Local Interpretable Model-Agnostic Explanations (LIME) for selected prediction instances of the proposed model. Each subfigure shows feature contributions to the predicted Bitcoin price, where green indicates positive impact and red indicates negative impact. Key features such as closing price, high, and low consistently exhibit strong influence, highlighting the model's instance-level interpretability

larly the closing price) are leading predictive performance, which further justifies the focus of the model on statistically relevant predictors. Second, the LIME explanations also hold local interpretable feature attributions, which, in turn, positively impacts trust in the model decision-making. By displaying the usefulness of recent lags of close, high, and low prices, the model portrays accuracy as well as interpretability. Cumulatively, these analyses make the proposed forecasting framework one that is not only accurate, however also can be explained and hence more favorable when considering practical applicability in financial decision-making systems where explainability is paramount. Although permutation importance and LIME can offer valuable insights into model behavior, both are post-hoc explainability methods that interpret the model once it has been trained as opposed to being inherently interpretable. As a result, they might not be a complete reflection of the internal decision-making procedure of deep learning models.

LIME, specifically, produces local explanations, by modeling behavior of models, by approximation around individual instances. These explanations may vary over time in time-series applications, like cryptocurrency forecasting, where a small perturbation in input data or sampling can result in varying interpretations over time windows. The consistency of the explanations may be influenced by this sensitivity.

Instead, permutation importance will evaluate feature relevance by measuring performance loss in the event of feature shuffling. Nonetheless, its outcomes may be affected by correlations of features and do not always mean causal relationships.

Intrinsic interpretability techniques, like attention in sequence models, give explanations as a component of the model structure, by indicating significant time steps or features learned in training. In spite of being more integratively interpretable, these methods tend to be computationally complex and were not within the scope of this study.

Thus, the interpretability findings of this study are to be considered as rough but informative information about the model behavior than as a set of causal explanations. Further investigation can be conducted in the future to include attention-based or inherently interpretable architectures to enhance explanation reliability further.

4.10 Potential Economic Implications

Although the main aim of this research is that of predictive accuracy, the findings might hold some possible meaning in the financial area. Even small gains in predicting performance can help in cryptocurrency markets, to create greater awareness of price fluctuation.

These predictive insights can aid in the decision-making processes, as they can assist analysts to make sense of the market trends and gauge the risks involved in very dynamic markets.

It should be noted, though, that the application of trading strategies, the modeling of transaction costs, and simulation of markets is not featured in this work. Hence, the results must be perceived as forecasting projections with the likely economic importance as opposed to financial recommendations.

These are additions to facilitate a better connection between predictive modeling and its possible relevance to computational economics without providing a further scope on the study.

4.11 Ablation Study

The ablation study examines thoroughly the contribution of each of the components of the proposed hybrid framework as shown in Table 6. First, pure LSTM and GRU models were implemented and achieved good baselines, however did not appear to capture long-term dependencies as well. The hybrid LIM-GRLST model enhanced forecasting with an improved accuracy in the forecast, as a result of the support of the deficiencies of each of these architectures. Moreover, the weighted hybrid model provided more consistent predictions whereas the neural hybrid introduced a trainable fusion mechanism resulting to better performance. Lastly, hyperparameter tuning also optimized the hybrid model with a significant reduction in the error metrics (MAE, MSE, RMSE) and an increase in the coefficient of determination (R^2). Such a development clearly shows that the suggested tuned hybrid model is more effective than all baselines, proving its usefulness in predicting cryptocurrency prices.

4.12 Comparison with State-of-the-art Studies

A comparison of recent research in cryptocurrency price forecasting is given in Table 7. Bold values show the best score. Deep learning architectures have been widely studied in the existing literature including LSTM, GRU and BiLSTM, frequently optimized by hyperparameter optimization and hybridization, and supplemented with auxiliary inputs like sentiment data. As an example, Ahmed et al. (2025) claims to have achieved good regression performance with BiLSTM and hyperparameter optimisation as well as feature extraction through autoencoders. In the same way, it was shown that Bayesian optimization is efficient in enhancing hybrid recurrent models (Kaur et al., 2025a), and the addition of sentiment and attention mechanisms (Seabe et al., 2025) is beneficial. Hybrid LSTMGRU models have also demonstrated bet-

Table 6 Ablation study showing performance related to every point

Model variant	Description	MAE	MSE	RMSE	R^2
Baseline LSTM	Standard LSTM for time series forecasting	0.0445	0.0055	0.0741	0.9277
Baseline GRU	Standard GRU for time series forecasting	0.0381	0.0048	0.0690	0.9373
Weighted hybrid	Weighted ensemble of LSTM and GRU outputs	0.0400	0.0049	0.0703	0.9348
Neural hybrid	Neural fusion layer for LIM-GRLST	0.0285	0.0027	0.0520	0.9644
Tuned hybrid	Hyperparameter optimized hybrid model	0.0209	0.0008	0.0279	0.9897

Table 7 Comparison of proposed model with existing studies

Ref.	Year	Model / Methodology	Reported metrics	Key observations
[35]	2025	BiLSTM, RNN, GRU with Autoencoder + Hyperband tuning	MSE: 0.001183, MAE: 0.02609, R^2: 0.9806	BiLSTM demonstrates strong regression performance
[36]	2024	LSTM with Bayesian-optimized RNN and BiLSTM	RMSE: ∼8% (relative)	Hybrid optimization improves predictive capability
[37]	2025	LSTM, GRU, BiLSTM with sentiment analysis and attention	Metric varies (reported in original study)	Sentiment-enhanced models improve forecasting behavior
[38]	2025	Hybrid LSTM–GRU	RMSE: 0.323, MAE: 0.464	Hybrid models outperform standalone architectures
Proposed	**2025**	**LSTM, GRU, Weighted Hybrid, Neural Hybrid, Tuned Hybrid (LIM-GRLST)**	**RMSE: 0.0279, MAE: 0.0209, R^2: 0.9897**	**Tuned hybrid achieves strong performance with interpretability**

Note: Reported results across studies are not directly comparable due to differences in datasets, preprocessing techniques, prediction horizons, and evaluation protocols. The table provides a contextual overview rather than a strict performance ranking

ter predictive performance compared to independent models (Omole & Enke, 2025; Raza et al., 2025).

Experimental results show that the tuned hybrid model proposed reveals a good performance in the realms of RMSE, MAE and R^2 in the controlled experimental context of the given study. Nevertheless, direct comparison to previous works must be taken with a grain of salt since variations in datasets, time periods, preprocessing methods, and measures of evaluation can have a bearing on the results being reported. Thus, the comparison was aimed at offering contextual information and not a strict standard. All models in this study were evaluated in the same experimental conditions in order to have a fair evaluation. Moreover, statistical significance of the identified performance improvements has been justified with the help of Diebold-Mariano test contributing to the strength and dependability of the suggested method.

4.13 Limitations and Future Research

Despite strong predictive performance, several limitations must be acknowledged. First, the study is based on Bitcoin daily price data (2013–2021), which may not fully capture evolving market dynamics, structural breaks, and regime shifts inherent in cryptocurrency markets. Second, the use of price-level data may introduce persistence effects, potentially inflating performance metrics such as R^2, and thus results should be interpreted with caution. Third, the evaluation is based on a fixed train–test split, which may not fully reflect real-world forecasting scenarios. Further-

more, the interpretability methods employed also have inherent limitations. LIME perturbs input features, which can disrupt temporal dependencies in time-series data, resulting in explanations that may not correspond to realistic sequences. Similarly, permutation importance breaks temporal order and feature correlations, limiting its ability to explain extreme events or nonlinear interactions. Therefore, these explanations should be considered indicative rather than definitive. In terms of generalization, the proposed model is trained on a single asset and a fixed historical period, and its applicability to other cryptocurrencies or future market conditions cannot be fully guaranteed. Although the hybrid architecture and hyperparameter optimization improve robustness, further validation across diverse datasets is required.

Future work will focus on several directions. First, advanced architectures such as attention mechanisms and transformer-based models can be integrated to further improve temporal modeling. Second, incorporating external factors such as social media sentiment and news data may enhance the model's ability to capture market volatility. Third, probabilistic and uncertainty-aware forecasting approaches can be explored to improve robustness under high volatility conditions. Additionally, future studies should adopt rolling window or walk-forward validation strategies and evaluate performance across multiple market regimes to ensure robustness. Finally, extending the framework to multiple cryptocurrencies and real-time streaming data, along with integrating trading strategies and economic validation, would improve its practical applicability in financial decision-making systems.

5 Conclusion

This study proposed a hybrid deep learning framework, LIM-GRLST, for cryptocurrency time-series forecasting by integrating LSTM and GRU architectures to capture both long-term dependencies and sequential patterns. Two hybrid strategies were explored: a neural hybrid model based on concatenation of LSTM and GRU representations, and a weighted hybrid model combining their predictions. Hyperparameter tuning further enhanced the performance of the neural hybrid model. Experimental results demonstrate that the proposed neural hybrid model outperforms standalone LSTM, GRU, and weighted hybrid approaches in terms of RMSE, MAE, and R^2. In addition, statistical validation using the Diebold–Mariano test confirms that the proposed model achieves statistically significant improvements over baseline models such as ARIMA and Naïve forecasts. To enhance model transparency, explainable AI techniques were incorporated. Permutation feature importance was used to identify globally influential features, while LIME provided local explanations for individual predictions. These analyses indicate that features such as closing price, high, and low values play a dominant role in forecasting outcomes, offering practical insights into model behavior.

Author Contributions WB conceptualization, data curation, writing manuscript. SR formal analysis, conceptualization, writing manuscript, supervision. MMI methodology, formal analysis, data curation. AR software, project administration, methodology. DRA funding acquisition, visualization, investigation. AGMM visualization, investigation, software. IA supervision, validation, writing and editing the manuscript. All authors reviewed the manuscript.

Funding This research was funded by the European University of Atlantic.

Availability of data and materials The dataset used in this study can be requested from the corresponding author.

Declarations

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Clinical Trial Number Not applicable.

Conflicts of Interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References

Aggarwal, A., Gupta, I., Garg, N., & Goel, A. (2019). Deep learning approach to determine the impact of socio economic factors on Bitcoin price prediction. In *2019 Twelfth International Conference on Contemporary Computing (IC3)* (pp. 1–5). https://doi.org/10.1109/IC3.2019.8844928

Ahmed, M., Adnan, M., Matar, A., Srayyih, F. H., Hammoodi, M., Mathil, T., & Jumaah, A. (2025). Predicting Bitcoin prices using deep learning. https://doi.org/10.2139/ssrn.5352108

Aljadani, A. (2022) $\mathrm{DLCP}^2\ \mathrm{F}$: a DL-based cryptocurrency price prediction framework. *Discover Artificial Intelligence*.

Alvi, J., Nizam, K., Jafri, S. M., & Rehan, M. (2024). Cryptocurrency predictive analytics: A comparative study of LSTM, CNN, and GRU models. *Pakistan Business Review*.

Amberkhani, A., Bolisetty, H., & Narasimhaiah, R. (2025). Revolutionizing cryptocurrency price prediction: Advanced insights from machine learning, deep learning and hybrid models. *Future Information*.

Andi, H. K. (2021). An accurate Bitcoin price prediction using logistic regression with LSTM machine learning model. *Journal of Soft Computing Paradigm*, *3*.

Ataei, S., Ataei, S. T., & Omidmand, P. (2025). Applications of deep learning to cryptocurrency trading: A systematic analysis. *Soft Computing*.

Birim, S. O., Comert, M., & Simsek, A. E. (2022). *An analysis for cryptocurrency price prediction using LSTM*. National and Global Fields: GRU and the bi-directional implications.

Das, J. D., Thulasiram, R. K., & Henry, C. (2024). Encoder-decoder based LSTM and GRU architectures for stocks and cryptocurrency prediction. *Journal of Risk and Financial Management*.

Dincer, E., & Kilimci, Z. H. (2025). From LSTM to GPT-2: Recurrent and transformer-based deep learning architectures for multivariate high-liquidity cryptocurrency price forecasting. *Symmetry, 18*(1), 32.

Farooq, A., Uddin, M. I., Adnan, M., Alarood, A. A., Alsolami, E., & Habibullah, S. (2024). Interpretable multi-horizon time series forecasting of cryptocurrencies by leveraging temporal fusion transformer. *Heliyon, 10*(22), 40142. https://doi.org/10.1016/j.heliyon.2024.e40142

Gur, Y. E. (2024). *Comparative analysis of deep learning models for silver price prediction: CNN, LSTM*. Akdeniz IIBF Journal: GRU and hybrid approach.

Hamayel, M. J., & Owda, A. Y. (2021). A novel cryptocurrency price prediction model using GRU, LSTM and Bi-LSTM machine learning algorithms. *AI, 2*(4), 477–496. https://doi.org/10.3390/ai2040030

Hasanli, R., & Dursun, M. (2026). A novel hybrid transformer-based deep learning approach for multi-step Bitcoin price forecasting. *Engineering, Technology & Applied Science Research, 16*(1), 32524–32533.

Hussein, N., & Abdulazeez, A. M. (2024). Bitcoin price prediction using hybrid LSTM-GRU models. *Indonesian Journal of Computer Science*.

Jiang, X. (2020). Bitcoin price prediction based on deep learning methods. *Journal of Mathematical Finance, 10*(1), 132–139. https://doi.org/10.4236/jmf.2020.101009

John, D. L., Binnewies, S., & Stantic, B. (2024). Cryptocurrency price prediction algorithms: A survey and future directions. *Forecasting, 6*(3), 637–671. https://doi.org/10.3390/forecast6030034

Kaur, R., Uppal, M., Gupta, D., Juneja, S., Arafat, S. Y., Rashid, J., Kim, J., & Alroobaea, R. (2025a). Development of a cryptocurrency price prediction model: Leveraging GRU and LSTM for Bitcoin, Litecoin and Ethereum. *PeerJ Computer Science*. Available at: https://europepmc.org/article/pmc/pmc11935774

Kaur, R., Uppal, M., Gupta, D., Juneja, S., & Arafat, S. Y. (2025b). *Development of a cryptocurrency price prediction model: leveraging GRU and LSTM for Bitcoin*. PeerJ Computer Science: Litecoin and Ethereum.

Kervancı, I. S., & Akay, M. F. (2020). Review on Bitcoin price prediction using machine learning and statistical methods. *Sakarya University Journal of Computer and Information Sciences, 3*, 177–187. https://doi.org/10.35377/saucis.03.03.774276

Kim, G., Shin, D. H., Choi, J. G., & Lim, S. (2022). A deep learning-based cryptocurrency price prediction model that uses on-chain data. IEEE.

Lee, M. C. (2025). Temporal fusion transformer-based trading strategy for multi-crypto assets using on-chain and technical indicators. *Systems, 13*(6), 474.

Lian, J. (2024). Comparative analysis of LSTM, GRU and transformer deep learning models for cryptocurrency ZEC price prediction performance. In *Proceedings of the international conference on financial innovation*.

Mazinani, A., Davoli, L., & Ferrari, G. (2025). Deep learning algorithms for cryptocurrency price prediction: A comparative analysis. *Distributed Ledger Technology*.

Murray, K., Rossi, A., Carraro, D., & Visentin, A. (2023). On forecasting cryptocurrency prices: A comparison of machine learning, deep learning, and ensembles. *Forecasting*.

Nair, M., Marie, M. I., & Abd-Elmegid, L. A. (2023). Prediction of cryptocurrency price using time series data and deep learning algorithms. *International Journal of Computer Science*.

Omole, O., & Enke, D. (2025). Using deep learning for Bitcoin price direction prediction: An empirical comparison of models and trading strategies. *Financial Innovation, 10*, 117. https://doi.org/10.1186/s40854-024-00643-1

Patel, N. P., Parekh, R., Thakkar, N., & Gupta, R. (2022). Fusion in cryptocurrency price prediction: A decade survey on recent advancements, architecture, and potential future directions. IEEE.

Peik, A., Chahooki, M. A. Z., Fard, A. M., & Sarram, M. A. (2025). Adaptive temporal fusion transformers for cryptocurrency price prediction. arXiv preprint.

Pintelas, E., Livieris, I. E., Stavroyiannis, S., Kotsilieris, T., & Pintelas, P. (2020). Investigating the problem of cryptocurrency price prediction: A deep learning approach. In *Artificial intelligence applications and innovations. IFIP Advances in information and communication technology* (vol. 584, pp. 99–110). Springer, cham. https://doi.org/10.1007/978-3-030-49186-4_9

Rajabi, S., Roozkhosh, P., & Farimani, N. M. (2022). MLP-based learnable window size for Bitcoin price prediction. *Applied Soft Computing*.

Rajkumar, S. (2025). Cryptocurrency Historical Prices: Bitcoin (CSV Dataset). https://www.kaggle.com/datasets/sudalairajkumar/cryptocurrencypricehistory?select=coin_Bitcoin.csv. Accessed Aug 2025

Raza, A., Rustam, F., Siddiqui, H. U. R., Flores, E. S., Mazón, J. L. V., Díez, T., Ripoll, I., Ashraf, M. A. V., & I. (2025). Ventilator pressure prediction employing voting regressor with time series data of patient breaths. *Health Informatics Journal, 31*(1), 14604582241295912.

Rizwan, M., Narejo, S., & Javed, M. (2019). Bitcoin price prediction using deep learning algorithm. In *2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS)* (pp. 1–7). https://doi.org/10.1109/MACS48846.2019.9024772

Seabe, P. L., Moutsinga, C. R. B., & Pindza, E. (2023). Forecasting cryptocurrency prices using LSTM, GRU, and bi-directional LSTM: a deep learning approach. *Fractal and Fractional*.

Seabe, R. L., Moutsinga, C. R. B., & Enter, E. (2025). Sentiment-driven cryptocurrency forecasting: analyzing LSTM, GRU, Bi-LSTM, and temporal attention model (TAM). *Social Network Analysis and Mining, 15*(1), 52.

Singh, A., Kumar, A., & Akhtar, Z. (2021). Bitcoin price prediction: A deep learning approach. In *2021 8th International conference on computing for sustainable global development (INDIACom)*. India: IEEE.

Ud Din, R., & Khan, S. H. (2026). TFT-ACB-XML: Decision-level integration of customized temporal fusion transformer and Attention-BiLSTM with XGBoost meta-learner for BTC price forecasting. arXiv preprint.

Zhang, S., Li, M., & Yan, C. (2022). The empirical analysis of Bitcoin price prediction based on deep learning integration method. *Computational Intelligence and Neuroscience, 2022*(1), 1265837. https://doi.org/10.1155/2022/1265837

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Wajeeha Badar[1] · Shabana Ramzan[1] · Ali Raza[2] · David Ramirez Alcocer[3,4,5,6] · Gerardo Mendez Mezquita[3,6,7,8] · Imran Ashraf[9]

✉ Shabana Ramzan
shabana@gscwu.edu.pk

✉ Imran Ashraf
imranashraf@yu.ac.kr

Wajeeha Badar
wajeehabadar1209@gmail.com

Ali Raza
ali.raza.scholarly@gmail.com

David Ramirez Alcocer
david.ramirez.alcocer@unini.edu.mx

Gerardo Mendez Mezquita
gerardo.mendez@unini.edu.mx

1 Department of Computer Science and IT, The Govt. Sadiq College Women University, Bahawalpur 63100, Pakistan

2 Department of Software Engineering, University of Lahore, 54000 Lahore, Pakistan

3 Universidad Europea del Atlantico, Isabel Torres 21, Santander 39011, Spain

4 Universidad Internacional Iberoamericana, Campeche 24560, Mexico

5 Fundacion Universitaria Internacional de Colombia, Bogota, Colombia

6 Universidad de La Romana, La Romana, República Dominicana

7 Universidad Internacional Iberoamericana Arecibo, Puerto Rico, PR 00613, USA

8 Universidade Internacional do Cuanza, Cuito, Bie, Angola

9 Department of Infromation and Communication Engineering, Yeungnam University, Gyeongsan-si 38541, Republic of Korea